

Adaptive schemes for incomplete quantum process tomography

Yong Siah Teo,^{1,2} Berthold-Georg Englert,^{1,3} Jaroslav Řeháček,⁴ and Zdeněk Hradil⁴

¹Centre for Quantum Technologies, National University of Singapore, Singapore 117543, Singapore

²NUS Graduate School for Integrative Sciences and Engineering, Singapore 117597, Singapore

³Department of Physics, National University of Singapore, Singapore 117542, Singapore

⁴Department of Optics, Palacky University, 17. listopadu 12, 77146 Olomouc, Czech Republic

(Dated: October 7, 2011)

We propose an iterative algorithm for incomplete quantum process tomography, with the help of quantum state estimation, based on the combined principles of maximum-likelihood and maximum-entropy. The algorithm yields a unique estimator for an unknown quantum process when one has less than a complete set of linearly independent measurement data to specify the quantum process uniquely. We apply this iterative algorithm adaptively in various situations and so optimize the amount of resources required to estimate the quantum process with incomplete data.

PACS numbers: 03.65.Ud, 03.65.Wj, 03.67.-a

I. INTRODUCTION

Quantum process tomography (QPT) is an important tool to characterize the operation of a given quantum channel [1–4]. Such a characterization is needed, for example, when one attempts to construct a quantum channel comprising multiple logic gates, each carrying out a specific quantum process. One such quantum channel for entanglement distillation, for instance, would consist of CNOT gates. A physical quantum process is described by a completely-positive map \mathcal{M} . That is, given a particular input quantum state ρ_i residing in the D_i -dimensional Hilbert space \mathcal{H} , the resulting output state ρ_o in the D_o -dimensional Hilbert space \mathcal{K} is given by

$$\rho_o = \mathcal{M}(\rho_i) = \sum_m K_m \rho_i K_m^\dagger, \quad (1)$$

with the Kraus operators K_m satisfying the relation $\sum_m K_m^\dagger K_m = 1_{\mathcal{K}}$. The K_m s are not unique and any other set of Kraus operators

$$K'_m = \sum_{m'} u_{m'm} K_{m'}, \quad (2)$$

where the $u_{m'm}$ s are the elements of a unitary matrix, also parameterizes the completely-positive map \mathcal{M} [5].

The idea behind QPT is to estimate such completely-positive maps with measurements. Much like quantum state tomography, the estimation of an unknown quantum process can be perceived as the estimation of a positive *Choi-Jamiólkowski operator* E_{true} that is represented by a $D_i D_o \times D_i D_o$ matrix [6]. Such an operator contains all accessible information about the quantum process. The standard QPT procedure involves the measurement of multiple copies of L different output states, with each output state corresponding to one of the L linearly independent input states $\rho_i^{(l)}$, thereby using a probability operator measurement (POM) of, say, M outcomes. The unknown operator E_{true} is estimated by linear-inversion of the LM measurement frequencies, which consists of

$D_i^2 D_o^2$ linearly independent constraints. Like the linear-inversion procedure for quantum state estimation, the resulting estimator obtained may not be positive. If that is the case, the estimator cannot be used for statistical predictions. This failure occurs whenever the observed relative frequencies of the measurement outcomes do not have consistent interpretation as probabilities. What is, therefore, called for, is an estimation procedure that ensures a physically meaningful estimator whatever the measurement data may be.

One statistically meaningful technique to obtain a positive estimator for E_{true} is the maximum-likelihood estimation procedure (ML) [7]. This can be applied to yield a unique estimator \hat{E}_{ML} as long as the measurement data obtained form a set of $D_i^2 D_o^2$ linearly independent constraints. We say that this set of measurement data is informationally complete. However, the number of linearly independent parameters increases rapidly with the dimensions and a complete characterization of E_{true} becomes unfeasible for complex processes. This is especially true when the quantum process acts on an infinite-dimensional Hilbert space [8]. The well-known method of *Direct Characterization of Quantum Dynamics* (DCQD) [2] was introduced to reduce the amount of measurement resources that are used for quantum process tomography. However, this method requires entangled input states and post-processing strategies that can be expensive when dealing with more complex quantum processes.

A more straightforward and conceptually different approach is to resort to informationally incomplete QPT. With this approach, less measurement resources are used to obtain an estimator for the unknown quantum process to a fair amount of accuracy. As a consequence, there exists a convex set of infinitely many ML estimators which are consistent with the measurement data. To choose the estimator which is least-biased from the convex set, we invoke the maximum-entropy principle [9] and choose the estimator with the largest entropy. Such an incomplete QPT can also give useful information about the quantum channel. In a typical tomography experiment, with data from measuring a finite number of copies, the

resulting quantum process estimator can never be exactly equal to E_{true} since experimental fluctuations are inevitable. One can only obtain an estimator that is close to E_{true} within a certain tomographic precision. Thus, MLME QPT is typically useful in providing a unique estimator for an unknown quantum process within a suitable tomographic precision using fewer incomplete measurement resources. As will be shown, this reduction in measurement resources is more pronounced for unitary quantum channels. Since E_{true} is unknown, one common practice is to gauge such a tomographic precision with another operator E_{prior} that is close to E_{true} , based on some prior information one has about the constructed quantum channel. The availability of such a E_{prior} for a given E_{true} will become useful and important in subsequent discussions.

The estimators obtained using the aforementioned method are least-biased with respect to the set of incomplete measurement data in the sense of the *entropy of the quantum process*. In Ref. [10], which is an analytical study of the conventional maximum-entropy method, the entropy functional for the Choi-Jamiólkowski operator E describing a quantum channel was introduced as $S(E) = -\text{tr}\{(E/D_i)\log(E/D_i)\}$ and this was shown to exhibit nice properties. In particular, this concave channel entropy functional has a unique maximum in E and is zero only when the quantum channel is unitary since E/D_i is then a rank-1 projector. However, the analytical results in [10] apply only to simple qubit channels and are difficult to extend to general quantum channels of greater complexity.

In this article, we shall extend the strategy in Ref. [11] and establish an adaptive iterative algorithm to search for the MLME estimator \hat{E}_{MLME} which maximizes both the likelihood and entropy functionals using the channel entropy functional in [10]. We first give some preliminary ideas on quantum process estimation in Sec. II. Then, in Sec. III, we will present the iterative MLME algorithm using variational principles to derive a steepest-ascent scheme and apply it to numerical simulations of two-qubit and three-qubit quantum channels. In Sec. IV, we will establish adaptive strategies to apply the MLME algorithm with the aim of minimizing the amount of measurement resources needed to perform incomplete QPT.

II. PRELIMINARIES OF QUANTUM PROCESS ESTIMATION

The estimation of the completely-positive map \mathcal{M} that describes an unknown quantum process, in the manner presented in Eq. (1), is isomorphic to the estimation of an unknown quantum state. This is a consequence of the well-known Choi-Jamiólkowski isomorphism [6, 7]. Let us define a maximally-entangled pure state $|\Psi_+\rangle = \sum_j |j\rangle_{\mathcal{H}} \otimes |j\rangle_{\mathcal{H}'} / \sqrt{D_i}$ in terms of the computational basis kets $|j\rangle_{\mathcal{H}} \otimes |j\rangle_{\mathcal{H}'}$. Here, the dimensions of the Hilbert spaces \mathcal{H} and \mathcal{H}' are both equal to the dimension D_i of the input Hilbert space. Using this basis, there exists a one-to-one correspondence between the map \mathcal{M} and a unique positive operator E defined as follows:

$$E \equiv D_i (\mathcal{I}_{\mathcal{H}} \otimes \mathcal{E}_{\mathcal{H}'})(|\Psi_+\rangle \langle \Psi_+|) \quad (3)$$

$$\equiv \sum_{jk} (|j\rangle \langle k|) \otimes \mathcal{M}(|j\rangle \langle k|), \quad (4)$$

with $\mathcal{I}_{\mathcal{H}}$ being the identity map. From Eq. (1), the alternative expression

$$E = \sum_m |\psi_m\rangle \langle \psi_m|, \quad (5)$$

with

$$|\psi_m\rangle = (1_{\mathcal{H}} \otimes K_m)|\Psi_+\rangle \sqrt{D_i}, \quad (6)$$

implies that the rank of E is equal to the number of linearly independent K_m s. It follows that E is rank-1 if the completely-positive map is described by a single unitary Kraus operator, and only then.

The output state can be expressed in terms of E by means of

$$\rho_o = \text{tr}_{\mathcal{H}} \{E(\rho_i^T \otimes 1_{\mathcal{K}})\}, \quad (7)$$

where the transposition is defined with respect to the computational basis. Hence, reconstructing the quantum process amounts to estimating the positive operator E . To do so, one requires a total of $D_i^2 D_o^2$ real parameters to specify the corresponding matrix. In the subsequent analyses, we shall consider trace-preserving maps, that is $\text{tr}\{\rho_i\} = \text{tr}\{\rho_o\}$ for any ρ_i , in which case the number of independent parameters is reduced to $D_i^2(D_o^2 - 1)$, with the constraints compactly written as

$$\text{tr}_{\mathcal{K}} \{E\} = 1_{\mathcal{H}}. \quad (8)$$

To estimate E , typically a set of L input states $\rho_i^{(l)}$, with N copies each, are sent through the quantum channel, one state at a time. The output state $\rho_o^{(l)}$ that corresponds to $\rho_i^{(l)}$ is measured with a POM consisting of M outcomes $\Pi_m \geq 0$ such that $\sum_m \Pi_m = 1_{\mathcal{K}}$. The probability of getting outcome Π_m for the input state $\rho_i^{(l)}$ is given by $p_{lm} = \text{tr}\{E(\rho_i^{(l)T} \otimes \Pi_m)\}/L$. Here, $p'_l \equiv \sum_m p_{lm} = 1/L$. If the LM parameters comprise $D_i^2 D_o^2$ linearly independent ones, the measurement data will be informationally complete. One can thus perform a complete quantum process estimation using the maximum-likelihood (ML) algorithm [7] and so obtain a unique positive estimator \hat{E}_{ML} by maximizing the likelihood functional

$$\mathcal{L}(E) = \prod_{l=1}^L \left(\prod_{m=1}^M p_{lm}^{n_{lm}} \right), \quad (9)$$

where the number of occurrences n_{lm} for the outcome Π_m obtained in an experiment with the input state $\rho_i^{(l)}$ are such that $n'_l \equiv \sum_m n_{lm} = N$.

In what follows, we shall discuss the situation in which one performs quantum process estimation with a set of informationally incomplete measurement data. All processes considered henceforth will be trace-preserving, but a generalization to those which are not trace-preserving is straightforward.

III. THE ITERATIVE ALGORITHM

We consider the optimization of the information functional

$$I(\lambda; E) = \lambda S(E) + \frac{1}{LN} \log \mathcal{L}(E), \quad (10)$$

where λ is a parameter which scales the entropy relative to the normalized log-likelihood and should be chosen with a very small value. When the measurement data are informationally complete, one sets λ to zero and optimizing $I(\lambda = 0; E)$ amounts to the ML problem [7]. In the same spirit as in [11], both our knowledge from the measurement data (contained in $\log(\mathcal{L}(E)) / LN$ which measures the information gain) and our ignorance (reflected in $S(E)$ which measures the lack of information) about the operator E are taken into account in such a way that our ignorance takes an infinitesimal weight. This introduces a small and smooth convex hill over the set of positive ML estimators which selects the one with the largest entropy. As in [11], the value of λ may be chosen such that both $\log(\mathcal{L}(E)) / LN$ and $S(E)$ remain almost constant with respect to λ .

To maximize $I(\lambda; E)$ with respect to E , we define the variation $E + \delta E = (1 + Z^\dagger)E(1 + Z)$, where Z is a small arbitrary operator such that Eq. (8) is satisfied, that is: $\text{tr}\{\delta E\} = 0$. Thus the most general expression for Z is

$$1+Z = (1 + \delta A) \left[\sqrt{\text{tr}_{\mathcal{K}} \{(1 + \delta A^\dagger) E (1 + \delta A)\}} \otimes 1_{\mathcal{K}} \right]^{-1}, \quad (11)$$

with an unrestricted infinitesimal δA . On the other hand, the variation of $I(\lambda; E)$ with respect to E gives $\text{tr}\{\delta E W\}$, where

$$W = \frac{1}{L} \sum_{lm} \frac{f_{lm}}{p_{lm}} \rho_i^{(l)T} \otimes \Pi_m - \frac{\lambda}{D_i} \left[1 + \log \left(\frac{E}{D_i} \right) \right] \quad (12)$$

and $f_{lm} = n_{lm}/LN$. Keeping only first-order variations systematically and imposing $\delta I(\lambda; E) > 0$, the method of steepest ascent leads us to

$$\delta A = \delta A^\dagger = \frac{\epsilon}{2} \left(W - \frac{1}{2} \text{tr}_{\mathcal{K}} \{WE + EW\} \otimes 1_{\mathcal{K}} \right) \quad (13)$$

for some small $\epsilon > 0$. Hence, to obtain the MLME estimator \hat{E}_{MLME} , one simply fixes $\lambda \ll 1$ and iterates the

equations

$$\begin{aligned} E_{n+1} &= (1 + Z_n^\dagger) E_n (1 + Z_n), \\ (\delta A)_n &= \frac{\epsilon}{2} \left(W_n - \frac{1}{2} \text{tr}_{\mathcal{K}} \{W_n E_n + E_n W_n\} \otimes 1_{\mathcal{K}} \right), \end{aligned} \quad (14)$$

where the expression for Z_n follows from Eq. (11) and W_n denotes the operator W in Eq. (12) evaluated for E_n . One may do so by starting from a randomly chosen operator E_0 and continue until the extremal equation for \hat{E}_{MLME} is satisfied with some pre-chosen numerical precision. To derive this extremal equation, we define the Lagrange functional [7]

$$\mathcal{D}(E) = I(\lambda; E) - \text{tr}\{\Lambda E\} \quad (15)$$

with the Lagrange operator $\Lambda \equiv h \otimes 1_K$ for the constraints in Eq. (8), where h is a Hermitian operator. Setting the variation of $\mathcal{D}(E)$ to zero gives the extremal equation

$$\Lambda \hat{E}_{\text{MLME}} \Lambda = W_{\text{MLME}} \hat{E}_{\text{MLME}} W_{\text{MLME}} \quad (16)$$

with $\Lambda = \sqrt{\text{tr}_{\mathcal{K}} \left\{ W_{\text{MLME}} \hat{E}_{\text{MLME}} W_{\text{MLME}} \right\}} \otimes 1_{\mathcal{K}}$.

Thus far, we have been assuming that the measurement outcomes Π_m give perfect detection of quantum systems. The iterative equations in Eq. (14) can be generalized to the case of imperfect detection. In this case it is clear that, if each of the M measurement outcomes Π_m is assigned a detection efficiency $\eta_m \leq 1$, one can always define a new set of M measurement outcomes $\tilde{\Pi}_m \equiv \eta_m \Pi_m$ such that $G \equiv \sum_m \tilde{\Pi}_m \neq 1_K$. The iteration procedure of Eq. (14) can still be used with the new set of POM outcomes $\tilde{\Pi}_m$ provided that the operator W in Eq. (12) is replaced by $W - W_0$, where

$$W_0 = \frac{1}{L \sum_{l'} p'_{l'}} \sum_l \rho_i^{(l) T} \otimes G \quad (17)$$

accounts for the copies that escape detection.

As an example, we apply the algorithm to numerical simulations on two-qubit channels, the CNOT gate described by the unitary operator

$$U_{\text{CNOT}} \hat{=} \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \end{pmatrix} \quad (18)$$

and a randomly generated non-unitary quantum channel described by a full-rank Choi-Jamiólkowski matrix, as well as the three-qubit Toffoli gate described by the unitary operator

To quantify the discrepancy between an MLME estimator and the true Choi-Jamiółkowski operator E_{true} , we use the trace-class distance

$$\mathcal{D}_{\text{tr}}(\hat{E}_{\text{MLME}}, E_{\text{true}}) = \frac{1}{2D_i} \text{tr}\left\{\left|\hat{E}_{\text{MLME}} - E_{\text{true}}\right|\right\}, \quad (20)$$

where $|A| = \sqrt{A^\dagger A}$ for any operator A . In these simulations, we take the D_i -dimensional projectors of a symmetrically informationally complete POM (SIC POM) as the input states [12]. One such set of states contains the minimal number of D_i^2 pure states $|\psi_j\rangle\langle\psi_j|$ such that

$$|\langle\psi_j|\psi_k\rangle|^2 = \frac{D_i\delta_{jk} + 1}{D_i + 1}. \quad (21)$$

As shown in Fig. 1, using the MLME algorithm for QPT can give fast convergence in terms of tomographic efficiency with a reduced number of input states as quantum resources. This reduction is especially significant for unitary processes, where the Choi-Jamiółkowski operators are rank-1. For nonunitary quantum processes described by matrices of larger rank, the tomographic efficiency will be lower as shown in the first plot of Fig. 1. This is expected in analogy with quantum state tomography where it is more difficult to reliably estimate highly-mixed states than nearly-pure ones.

IV. ADAPTIVE STRATEGIES

An interesting question to ask with regard to incomplete QPT is whether one can perform it in an optimal way given the available resources by means of adaptive strategies. Here optimality refers to the minimization of the amount of resources (input states or measurements) used to perform incomplete QPT such that the distance between \hat{E}_{MLME} and E_{true} reaches a certain desired value. Very frequently, despite the fact that E_{true} is always unknown, one has a rough idea of an operator E_{prior} which may be close to E_{true} based on some prior information about the unknown E_{true} . This scenario is reasonable and typical when one designs a quantum channel experimentally which performs an expected quantum operation, with errors arising from imperfections of the components that make up the channel. We shall establish adaptive strategies which make use of such an operator in order to select, with the help of the MLME algorithm, resources for incomplete QPT in an optimal way. We refer to such tomography schemes as the *adaptive MLME quantum process tomography* (AMLME QPT) schemes.

We will focus on adaptive strategies to choose the input states optimally. This can be reviewed in two separate cases: The case in which a fixed set of linearly independent input states is used (Sec. IV A) and that in which arbitrary input states can be generated for incomplete QPT (Sec. IV B). The optimization of the POM will be studied on a later occasion.

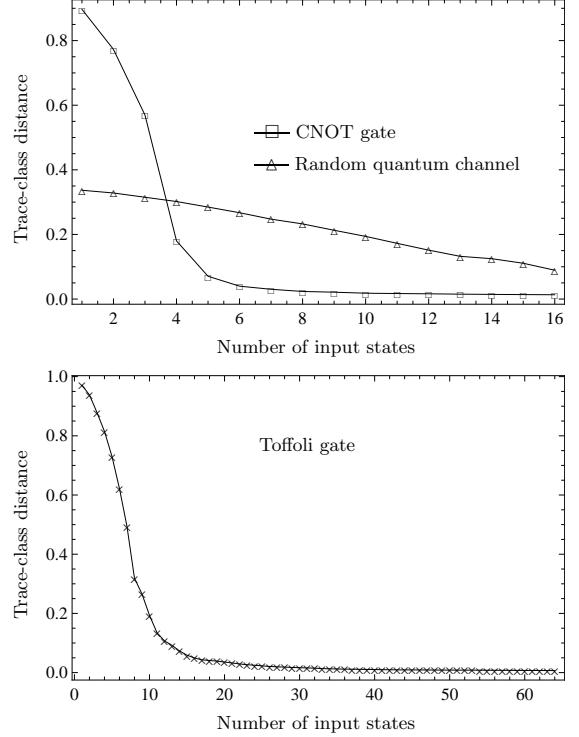


FIG. 1. Numerical simulations on the two-qubit ($d = 2^2$) and three-qubit ($d = 2^3$) quantum channels where $D_i = D_o = d$. The projectors of symmetric informationally complete POMs (SIC POMs) are chosen as the linearly independent input states for all the simulations ($L = d^2$). For the measurements, informationally complete POMs consisting of tensor products of qubit SIC POMs are used ($M = d^2$). Each qubit SIC POM consists of a set of pure states whose Bloch vectors form the “legs of a tetrahedron” in the Bloch sphere. For the two-qubit channels, $N = 10^4$ and an average over 50 experiments is taken to compute the trace-class distances. For the three-qubit channel, the measurement data are generated without statistical noise. For unitary channels, one can see that the MLME algorithm can still give fairly accurate estimations with a smaller number of input states than that of a linearly independent set. Numerical simulations of arbitrary two-qubit and three-qubit unitary channels suggest that the number is approximately $d^2/2$ for SIC POM input states, above which there is insignificant tomographic improvement.

A. Optimization over a fixed set of linearly independent input states

In the previous section, we considered the projectors of the SIC POMs, which are known to have optimal tomographic efficiencies, as input states in the numerical simulations. Since these POMs are symmetric in the sense of Eq. (21), any ordering of the input states in a given set gives the same plots in Fig. 1. In practice, however, such entangled states are difficult to produce and one typically has access to a set of separable states [13] for measurements instead. In this case, there no longer exists such a symmetry and the tomographic performance depends on

the order of the input states chosen, possibly strongly so. We propose to optimize the tomographic performance by choosing the input states adaptively based on the measurement data collected from the previously chosen input states, thereby using the prior E_{prior} .

To describe the adaptive strategy, let us consider a set of $L \geq D_i^2$ input states in which D_i^2 of them are linearly independent. Suppose that N , which is a fixed integer for all input states, copies of a randomly chosen input state $\rho_i^{(1)}$ are sent through the quantum channel and the first set of measurement data $\{\nu_{11}, \dots, \nu_{1M}\}$, $\sum_m \nu_{1m} = 1$, is collected. With these data $f_{1m} \equiv \nu_{1m}$, one obtains the first MLME estimator $\hat{E}_{\text{MLME}}^{(1)}$. To select the next input state out of the remaining $L - 1$ states, we take E_{prior} as a gauge for E_{true} to generate $L - 1$ sets of probabilities respectively from the $L - 1$ states. Each set of probabilities is then treated as the set of frequencies $\{\nu_{21}^{(k)}, \dots, \nu_{2M}^{(k)}\}$, for the corresponding input state k . Hence, one has $L - 1$ sets of measurement data, each set being the combined data $\{\nu_{11}, \dots, \nu_{1M}, \nu_{21}^{(k)}, \dots, \nu_{2M}^{(k)}\}/2$ with the normalized frequencies $f_{1m} \equiv \nu_{1m}/2$ and $f_{2m}^{(k)} \equiv \nu_{2m}^{(k)}/2$ such that $\sum_m (f_{1m} + f_{2m}^{(k)}) = 1$ for each k , and the corresponding $L - 1$ projected MLME estimators $\hat{E}_{\text{MLME},k}^{(2)}$.

The value of k is selected such that a chosen figure of merit which quantifies the distance between $\hat{E}_{\text{MLME},k}^{(2)}$ and $\hat{E}_{\text{MLME}}^{(1)}$ is the largest, so that there is a high chance for the next MLME estimator to be closer to E_{true} . As an example, the figure of merit is taken to be the trace-class distance $\mathcal{D}_{\text{tr}}(\hat{E}_{\text{MLME},k}^{(2)}, \hat{E}_{\text{MLME}}^{(1)})$.

With this input state, the second estimator $\hat{E}_{\text{MLME}}^{(2)}$ is then obtained with MLME QPT. One repeats this procedure for subsequent input states until the distance $\mathcal{D}_{\text{tr}}(\hat{E}_{\text{MLME}}^{(l+1)}, \hat{E}_{\text{MLME}}^{(l)})$ is below some preset threshold. An alternative to this would be to minimize the trace-class distance $\mathcal{D}_{\text{tr}}(\hat{E}_{\text{MLME},k}^{(l+1)}, E_{\text{prior}})$.

It is important to understand that in this strategy, the prior information E_{prior} is *not* used to reconstruct the unknown quantum process in any way. It serves only as a means to optimally select the input states from the given set so as to maximize the tomographic convergence. This adaptive strategy also relies partially on the measurement data obtained in the experiment. We have thus introduced an operational method of using the prior information to minimize the amount of resources needed to perform reliable MLME QPT without introducing any artifacts coming from the prior information into the reconstruction procedure. To summarize, the adaptive MLME strategy is as follows:

1. Randomly choose $\rho_i^{(1)}$ from the set of L input states and set $l = 1$.
 - (a) Perform QPT using $\rho_i^{(l)}$ and obtain the set of frequencies $\{\nu_{l1}, \dots, \nu_{lM}\}$, $\sum_m \nu_{lm} = 1$.
 - (b) Set $\nu = \bigcup_{j=1}^l \{\nu_{j1}, \dots, \nu_{jM}\}/l$.

- (c) Invoke the MLME algorithm with ν and obtain $\hat{E}_{\text{MLME}}^{(l)}$. Use E_{prior} to compute the frequencies $\{\nu_{l+11}^{(k)}, \dots, \nu_{l+1M}^{(k)}\}$, $\sum_m \nu_{l+1m}^{(k)} = 1$, from the remaining input states, with k labeling the remaining $L - l$ states.
- (d) Define $L - l$ sets of accumulated frequencies $(\nu \cup \{\nu_{l+11}^{(k)}, \dots, \nu_{l+1M}^{(k)}\})/(l + 1)$ and calculate the $L - l$ projected MLME estimators $\hat{E}_{\text{MLME},k}^{(l+1)}$.
- (e) Set $\rho_i^{(l+1)}$ as the input state corresponding to k such that $\mathcal{D}_{\text{tr}}(\hat{E}_{\text{MLME},k}^{(l+1)}, \hat{E}_{\text{MLME}}^{(l)})$ is largest.

2. Set $l = l + 1$ and repeat Steps 1(a)–1(e).

B. Optimization over the Hilbert space

More generally, the adaptive strategy may be extended to the case in which one has access to the entire Hilbert space of states. In other words, the task is to search for the next optimal input state $\rho_i^{(L+1)}$ from the D_i -dimensional Hilbert space based on the measurement data ν_{lm} obtained in the experiment from L previously chosen input states, where $\sum_m \nu_{lm} = 1$ for all l , and the prior information E_{prior} about the unknown quantum process.

To this end, we define the normalized *projected* log-likelihood functional

$$\log \tilde{\mathcal{L}}(E, \rho) = \sum_{lm} \frac{\nu_{lm}}{L+1} \log(\tilde{p}_{lm}) + \sum_m \frac{\tilde{\nu}_m}{L+1} \log(\tilde{p}_m), \quad (22)$$

where

$$\tilde{p}_{lm} = \text{tr} \left\{ E \frac{\rho_i^{(l)T} \otimes \Pi_m}{L+1} \right\},$$

$$\tilde{\nu}_m = \text{tr} \{ E_{\text{prior}} \rho^T \otimes \Pi_m \} \text{ and } \tilde{p}_m = \text{tr} \left\{ E \frac{\rho^T \otimes \Pi_m}{L+1} \right\}$$

with l always running from 1 to L over all previously used input state labels. This projected log-likelihood functional is a good approximation to the log-likelihood functional for the situation in which the state ρ is chosen as the next input state for the experiment as long as E_{prior} is not too far away from E_{true} . The projected frequencies $\tilde{\nu}_m$ estimate the actual frequencies one gets when measuring the input state ρ . An optimal input state $\rho_i^{(L+1)}$ and the corresponding Choi-Jamiołkowski operator are chosen as the positive estimators that maximize this projected log-likelihood functional.

Coincidentally, this maximum projected log-likelihood (MPL) procedure is equivalent to minimizing the cross

entropy functional $\mathcal{C}(E, \rho) = -\log \tilde{\mathcal{L}}(E, \rho)$ [14, 15] over all positive operators subjected to the respective constraints for ρ and E . Hence, another way of understanding this procedure is to first regard both the incomplete measurement data collected after using L input states and E_{prior} as the full prior knowledge one has about the unknown E_{true} . The statistical motivation for MPL or minimizing $\mathcal{C}(E, \rho)$ is, loosely speaking, to obtain estimators which are as compatible with this prior knowledge as possible by minimizing the entropy of the prior knowledge $\mathcal{C}(E, \rho)$. We will provide some more arguments related to this optimization technique in the later part of this section.

To carry out the optimization, we consider the response of $\log \tilde{\mathcal{L}}(E, \rho)$ to small variations of both ρ and E . After some similar calculations as in Sec. III, we obtain the MPL iterative equations

$$\begin{aligned} E_{n+1} &= (1 + Z_n^\dagger) E_n (1 + Z_n), \\ \rho_{n+1} &= \frac{(1 + \epsilon_2 \Xi_n) \rho_n (1 + \epsilon_2 \Xi_n)}{\text{tr}_{\mathcal{H}} \{(1 + \epsilon_2 \Xi_n) \rho_n (1 + \epsilon_2 \Xi_n)\}}, \end{aligned} \quad (23)$$

where Z_n is defined by Eq. (11) with

$$\begin{aligned} (\delta A)_n &= \frac{\epsilon_1}{2} \left(X_n - \frac{1}{2} \text{tr}_{\mathcal{K}} \{X_n E_n + E_n X_n\} \otimes 1_{\mathcal{K}} \right), \\ X_n &= \sum_{lm} \frac{\nu_{lm} \rho_i^{(l)T} \otimes \Pi_m}{(L+1)^2} + \sum_m \frac{\tilde{\nu}_m \rho^T \otimes \Pi_m}{(L+1)^2}, \end{aligned} \quad (24)$$

and

$$\begin{aligned} \Xi_n &\equiv Y_n - \text{tr}_{\mathcal{H}} \{Y_n \rho_n\}, \\ Y_n &= \text{tr}_{\mathcal{K}} \left\{ \left[\sum_m \frac{1_{\mathcal{H}} \otimes \Pi_m}{L+1} \right. \right. \\ &\quad \times \left. \left. \left(\log(\tilde{p}_m) E_{\text{prior}} + \frac{\tilde{\nu}_m}{(L+1)\tilde{p}_m} E \right) \right]^T \right\}. \end{aligned} \quad (25)$$

The MPL estimators satisfy the extremal equations

$$\begin{aligned} \tilde{\Lambda} \hat{E}_{\text{MPL}} \tilde{\Lambda} &= X_{\text{MPL}} \hat{E}_{\text{MPL}} X_{\text{MPL}}, \\ \hat{\rho}_{\text{MPL}} Y_{\text{MPL}} &= Y_{\text{MPL}} \hat{\rho}_{\text{MPL}} = \text{tr}_{\mathcal{H}} \{Y_{\text{MPL}} \hat{\rho}_{\text{MPL}}\} \hat{\rho}_{\text{MPL}}, \end{aligned} \quad (26)$$

where

$$\tilde{\Lambda} = \sqrt{\text{tr}_{\mathcal{K}} \{X_{\text{MPL}} \hat{E}_{\text{MPL}} X_{\text{MPL}}\}} \otimes 1_{\mathcal{K}}. \quad (27)$$

The small parameters ϵ_1 and ϵ_2 are positive numbers. Thus, to carry out the MPL procedure, one iterates Eqs. (23) until Eqs. (26) are satisfied with a preset numerical precision.

There is one important feature of this optimization scheme. From Eq. (22), we note that $\log \tilde{\mathcal{L}}(E, \rho)$ is neither convex nor concave in ρ and hence can have multiple local maxima. Thus, the MPL optimization is non-convex. To generate these local-maxima estimators, one

can start from multiple randomly chosen starting points and perform the iterations. Thereafter, the state estimator $\hat{\rho}_{\text{MPL}}$ to be chosen as the next input state $\rho_i^{(L+1)}$ is such that its corresponding \hat{E}_{MPL} gives the largest trace-class distance away from the previous MLME estimator $\hat{E}_{\text{MLME}}^{(L)}$, which is obtained from the data of the previously chosen L input states, over all generated pairs of MPL estimators $(\hat{\rho}_{\text{MPL}}, \hat{E}_{\text{MPL}})$. Again, one may also minimize the distance between \hat{E}_{MPL} and E_{prior} . Let us summarize the adaptive MPL-MLME strategy with the following scheme:

1. Randomly choose $\rho_i^{(1)}$ as the first input state and set $l = 1$.
 - (a) Perform QPT using $\rho_i^{(l)}$ and obtain the set of frequencies $\{\nu_{l1}, \dots, \nu_{lM}\}$, $\sum_m \nu_{lm} = 1$.
 - (b) Set $\nu = \bigcup_{j=1}^l \{\nu_{j1}, \dots, \nu_{jM}\} / l$.
 - (c) Invoke the MLME algorithm with ν and obtain $\hat{E}_{\text{MLME}}^{(l)}$.
 - (d) Using E_{prior} , generate a set of pairs of MPL estimators $(\hat{\rho}_{\text{MPL}}, \hat{E}_{\text{MPL}})$, where the states $\hat{\rho}_{\text{MPL}}$ were not part of the l input states previously used, by iterating Eqs. (26) from different, randomly chosen starting points.
 - (e) Set $\rho_i^{(l+1)}$ as the input state corresponding to the state estimator $\hat{\rho}_{\text{MPL}}$ such that $\mathcal{D}_{\text{tr}}(\hat{E}_{\text{MPL}}, \hat{E}_{\text{MLME}}^{(l)})$ is the largest.
2. Set $l = l + 1$ and repeat Steps 1(a)–1(e).

With this, let us first compare the performances of the three proposed schemes, namely the non-adaptive MLME scheme in Sec. III, the adaptive MLME scheme in Sec. IV A and the adaptive MPL-MLME scheme. For this purpose, we consider two quantum processes, the first being an imperfect CNOT gate whose action is described by the Kraus operators

$$K_1 = \sqrt{1-\epsilon} U_{\text{CNOT}} \quad \text{and} \quad K_2 = \sqrt{\epsilon}. \quad (28)$$

This first channel is a CNOT gate with probability $1 - \epsilon$ and does nothing to the input states with probability ϵ , an imperfect CNOT gate represented by a rank-2 Choi-Jamiókowski operator. The second process is described by the Kraus operators

$$K_1 = \sqrt{1-\epsilon} U_{\text{CNOT}} \quad \text{and} \quad \{K_j = \sqrt{\epsilon} B_j\}_{j=2}^{16}, \quad (29)$$

where the 15 operators B_j are randomly generated and satisfy the equation $\sum_j B_j^\dagger B_j = 1_{\mathcal{K}}$. This second channel, which is represented by a full-rank matrix, is a CNOT gate with probability $1 - \epsilon$ and randomly perturbs the input states with probability ϵ due to additional noise. As an example, we set $\epsilon = 0.1$. Figure 2 shows the numerical results.

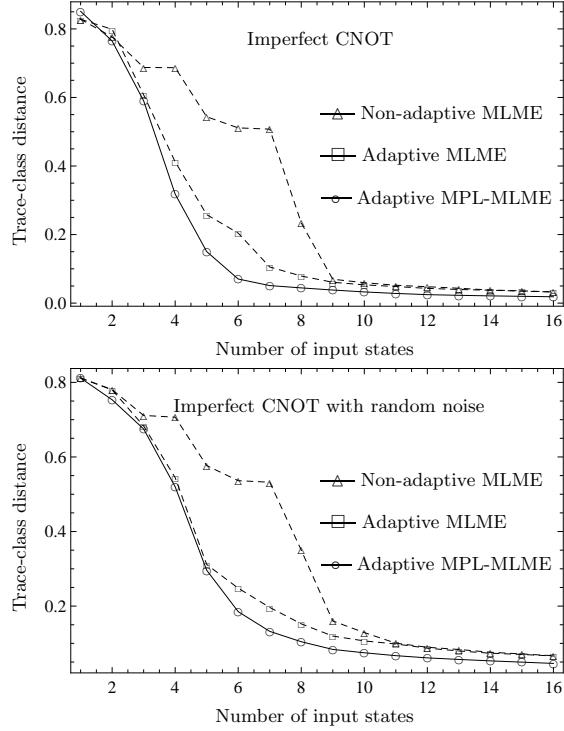


FIG. 2. A comparison of three incomplete QPT schemes: the non-adaptive MLME scheme, the adaptive MLME scheme and the adaptive MPL-MLME scheme. Monte Carlo simulations are carried out on two different types of imperfect CNOT gates described in the text. Here, $N = 10^4$ and an average over 50 experiments is taken to compute the trace-class distances. For both the non-adaptive as well as the adaptive MLME schemes, the 16 linearly independent input states are chosen to be tensor products of projectors of the kets $|0\rangle$, $|1\rangle$, $(|0\rangle + |1\rangle)/\sqrt{2}$ and $(|0\rangle + |1\rangle i)/\sqrt{2}$ [16]. For all schemes, the POM outcomes are chosen to be the tensor products of qubit SIC POMs. The tomographic performance of the adaptive MPL-MLME scheme is the best out of the three. The plots show that the tomographic efficiency can be further improved by optimizing the input states over the Hilbert space instead of restricting to a fixed set of linearly independent input states, albeit the small difference in tomographic performance between the two adaptive schemes for some quantum processes.

Next, to understand how this adaptive MPL-MLME strategy can lead to an optimization in tomographic performance, we need to know how increasing the number of input states used in AMLME QPT can affect the corresponding MLME estimators. Since we are considering only a subset of the full linearly independent input states in general, there exists a convex set of estimators \hat{E}_{ML} maximizing the likelihood functional for a given set of informationally incomplete measurement data. This means that the likelihood functional possesses a plateau hovering over this convex set of estimators. As the number of input states L used increases, the likelihood plateau will either remain unchanged (if no additional information about E_{true} is gained after performing QPT with new

input states) or decrease in size (if new independent information is obtained). Thus in general, the plateau will continue to shrink to a point when a full set of linearly independent input states is used.

We conjecture that the adaptive MPL-MLME strategy optimizes the rate of decrease in the size of the likelihood plateau by maximizing the normalized projected log-likelihood functional with respect to the next input state. A point of view to justify this conjecture is to interpret the maximum of the normalized log-likelihood functional $\log(\mathcal{L}(E))/LN$ as the maximum information gain from the measurement data. When the number of copies N is infinite, the data are noiseless and the resulting maximum information gain is $\sum_{lm} f_{lm} \log(f_{lm})$, which is the negative of the Shannon entropy of the measurement data. For finite N , the maximum information gain over the space of density operators will typically be lower than the true maximum due to the positivity constraint, especially when E_{true} is highly rank-deficient. In this language, the MPL-MLME strategy attempts to maximize this maximum information gain as much as possible via the optimization of future input states over the entire Hilbert space of density operators, using the normalized projected log-likelihood functional as an estimate for the actual normalized log-likelihood functional describing future measurements. This is a possible explanation for the optimal decrease in the likelihood plateau size since one has maximal knowledge about the unknown E_{true} gained with the optimized input states and so the ambiguity in the estimators is minimized.

We illustrate this point by considering the imperfect CNOT gate with $\epsilon = 0.1$ described by Eq. (28). Since the boundary of the likelihood plateau is complicated, we shall estimate its size numerically by first generating $N_0 = 10^3$ ML estimators $\hat{E}_{\text{ML}}^{(j)}$ labeled with the index j for a given set of measurement data. Next, in the same spirit as in numerical sampling, we can define the operator centroid

$$\bar{E}_{\text{ML}} = \frac{1}{N_0} \sum_{j=1}^{N_0} \hat{E}_{\text{ML}}^{(j)} \quad (30)$$

for this generated set of estimators and the normalized Hilbert-Schmidt standard deviation

$$\Delta = \frac{1}{D_i} \sqrt{\frac{\sum_{j=1}^{N_0} \text{tr} \left\{ \left(\hat{E}_{\text{ML}}^{(j)} - \bar{E}_{\text{ML}} \right)^2 \right\}}{2N_0}} \quad (31)$$

away from the centroid. Thus, $0 \leq \Delta \leq 1$. For sufficiently large N_0 , the size of the plateau may be well approximated by the spread Δ . Figure 3 compares the respective performances of the the three proposed schemes by analyzing the size of the likelihood plateau and the maximum of the normalized log-likelihood functional. From Fig. 3, it is crucial to understand that Δ does not, strictly speaking, decrease monotonically with increasing height of the normalized log-likelihood functional. A

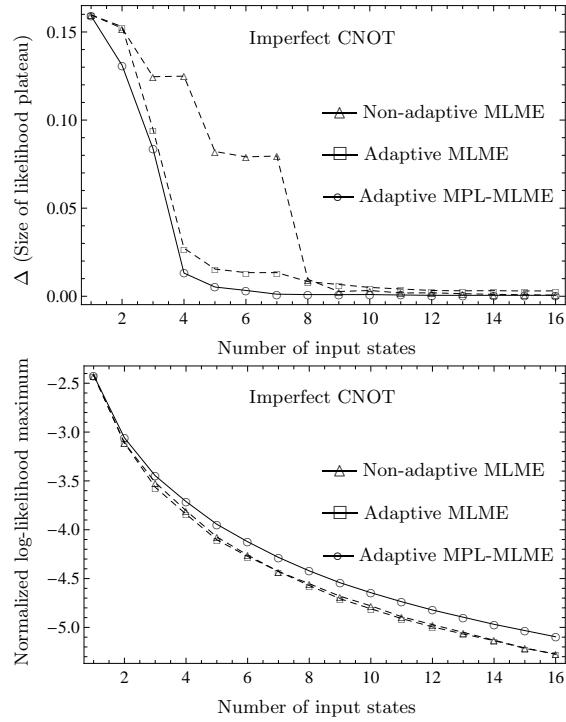


FIG. 3. The dependence of the size of the likelihood plateau (Δ) and the normalized log-likelihood maximum on the number of input states. The respective performances of the non-adaptive MLME scheme, the adaptive MLME scheme and the adaptive MPL-MLME scheme are computed based on noiseless measurement data for an imperfect CNOT gate with $\epsilon = 0.1$. For both the non-adaptive MLME scheme and the adaptive MLME scheme, the 16 linearly independent input states are chosen to be tensor products of projectors of the kets $|0\rangle$, $|1\rangle$, $(|0\rangle + |1\rangle)/\sqrt{2}$ and $(|0\rangle + |1\rangle i)/\sqrt{2}$. For all schemes, the POM outcomes are chosen to be the tensor products of qubit SIC POMs. From the plot, the rate of decrease of Δ is the greatest with the adaptive MPL-MLME scheme. The increase in the normalized log-likelihood maxima with the adaptive MPL-MLME scheme may also be interpreted as greater maximum information gain after measurements using the optimal input states as compared to the other schemes.

counterexample is shown in the figure, that is a significant decrease in Δ for the adaptive MLME scheme as compared to the non-adaptive one with the corresponding slight decrease in its normalized log-likelihood maxima. We emphasize that what the adaptive MPL-MLME strategy exploits is the possible *trend* of this behavior.

To end this part of the section, we comment that the aforementioned idea can be applied to adaptively choose the next set of POM outcomes Π_j based on the collected measurement data. However, to perform the optimization successfully requires the solutions to more technical problems which include ensuring that the POM outcomes are linearly independent after the optimization. This project is left for future studies.

C. A combination of both adaptive strategies

Let us begin this final part of the section by reviewing the non-convex feature of the MPL-MLME strategy discussed in Sec. IV B. The presence of multiple local-maxima estimators which are linearly independent is an important element of the MPL-MLME strategy as it provides linearly independent input states which are optimal for measurement based on the data obtained from the experiments. In general, because of the non-linearity of Eq. (26), it is difficult to determine the number of such linearly independent extremal solutions for a given set of measurement data by analytical means. One can only search for as many linearly independent local-maxima estimators $\hat{\rho}_{\text{MPL}}$ as possible via numerical optimizations from different starting points within a reasonable time period.

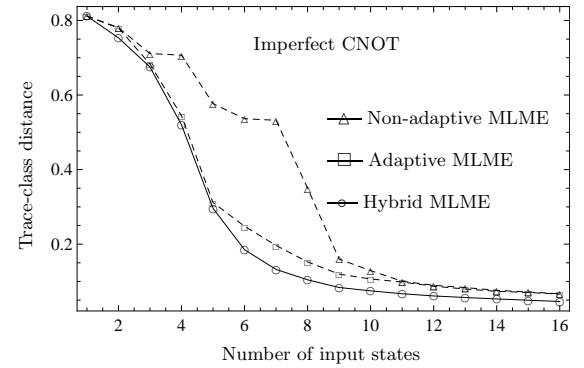


FIG. 4. A comparison of three incomplete QPT schemes: the non-adaptive MLME scheme, the adaptive MLME scheme and a combination of the adaptive MPL-MLME scheme and the adaptive MLME scheme (hybrid scheme). Monte Carlo simulations are carried out on the imperfect CNOT gate with $\epsilon = 0.1$. Here, $N = 10^4$ and an average over 50 experiments is taken to compute the trace-class distances. For both the non-adaptive as well as the adaptive MLME schemes, the default set of 16 linearly independent input states are chosen to be tensor products of projectors of the kets $|0\rangle$, $|1\rangle$, $(|0\rangle + |1\rangle)/\sqrt{2}$ and $(|0\rangle + |1\rangle i)/\sqrt{2}$. For all schemes, a set of 16 randomly generated positive operators, which are all linearly independent of one another, are used to form the POM. For this POM, the average repetition frequency of the adaptive MPL-MLME scheme is very high after four input states are used. The first input state for all schemes is chosen to be the same separable state $\rho_i^{(1)} = |00\rangle\langle 00|$. For the third scheme, the second to the fourth input states (shaded region) are optimized using the adaptive MPL-MLME strategy and the subsequent input states are chosen via the adaptive MLME strategy using the default set of input states which excludes $|00\rangle\langle 00|$. The plot shows that the overall performance of the combined strategy is better than the adaptive MLME strategy alone.

Another technical subtlety is that these local-maxima estimators tend to repeat themselves during the optimization. Hence, a local-maxima estimator which was chosen as one of the input states earlier may reappear in later optimizations. The repetition frequency strongly

depends on the POM chosen to measure the output states. The examples given thus far make use of the product tetrahedron measurements as the POM and the resulting MPL optimizations give linearly independent estimators with few repetitions. This may not be the case for other types of POM. In view of this, another way of doing AMLME QPT is to use both adaptive strategies in Secs. IV A and IV B interchangeably, the *hybrid* MLME strategy. For example, one can start with the adaptive MPL-MLME strategy for tomography and when the repetition rate increases as more input states have been used, one may switch to the first adaptive MLME strategy. Figure 4 suggests that such a hybrid MLME strategy can further improve the tomographic performance as compared with the adaptive MLME strategy alone.

V. SUMMARY

We have established adaptive numerical strategies to perform incomplete quantum process tomography. One may choose whichever strategy that is convenient to her to carry out tomography depending on the available types of measurement resources at hand. Each of these strategies combines the simplicity of incomplete quantum process tomography using quantum state estimation with good tomographic performances using optimization techniques. It can never be overemphasized that, although some prior information is necessary for each adaptive strategy, such information is *never* used in the estima-

tion of the unknown quantum process. Rather, the prior information is utilized to adaptively select future measurement resources, the input states in our context, based on the current measurement data, to optimize the tomographic performance. The discussions presented in this article, therefore, provide a means of obtaining estimators for the unknown quantum process using incomplete resources which are typically within reasonably good experimental precisions. These estimators are statistically meaningful in that they are least-biased with respect to a set of informationally incomplete measurement data and are hence suitable for partial characterization of quantum processes. This is in contrast with the standard quantum process tomography which generally requires a huge amount of informationally complete measurement resources.

ACKNOWLEDGMENTS

We would like to thank Guo Chuan Thiang and Huangjun Zhu for fruitful and stimulating discussions. This work is supported by the NUS Graduate School for Integrative Sciences and Engineering and the Centre for Quantum Technologies which is a Research Centre of Excellence funded by Ministry of Education and National Research Foundation of Singapore, as well as the Czech Ministry of Education, Project MSM6198959213, and the Czech Ministry of Industry and Trade, Project FR-TI1/364.

-
- [1] Throughout this article, the words “quantum process” and “quantum channel” will be used interchangeably.
 - [2] M. Mohseni, A. T. Rezakhani, and D. A. Lidar, Phys. Rev. A **77**, 032322 (2008).
 - [3] J. L. O’Brien, G. J. Pryde, A. Gilchrist, D. F.V. James, N. K. Langford, T. C. Ralph, and A. G. White, Phys. Rev. Lett. **93**, 8 (2004).
 - [4] J. F. Poyatos, J. I. Cirac, and P. Zoller, Phys. Rev. Lett. **78**, 2 (1997).
 - [5] M. A. Nielsen and I. L. Chuang, *Quantum Computation and Quantum Information*, Cambridge University Press, Cambridge, (2000).
 - [6] M. Choi, Linear Algebr. Appl. **10**, 285 (1975); A. Jamiołkowski, Rep. Math. Phys. **3**, 275 (1972).
 - [7] M. Paris, and J. Řeháček, *Lecture Notes in Physics 649 - Quantum State Estimation*, Springer, Berlin Heidelberg (2004); J. Fiurášek and Z. Hradil Phys. Rev. A **63** 020101(R) (2001).
 - [8] S. Rahimi-Keshari, A. Scherer, A. Mann, A. T. Rezakhani, A. I. Lvovsky, and B. C. Sanders, New J. Phys. **13**, 013006 (2011); M. Lobino, D. Korystov, C. Kupchak, E. Figueroa, B. C. Sanders, and A. I. Lvovsky, Science **322**, 563 (2008).
 - [9] E. T. Jaynes, Phys. Rev. **106**, 620 (1957), Phys. Rev. **108**, 171 (1957).
 - [10] M. Ziman, Phys. Rev. A **78**, 032118 (2008).
 - [11] Y. S. Teo, H. Zhu, B.-G. Englert, J. Řeháček, and Z. Hradil, Phys. Rev. Lett. **107**, 020404 (2011).
 - [12] A. J. Scott, and M. Grassl, J. Math. Phys. **51**, 042203 (2010); A. Ling, A. Lamas-Linares, and C. Kurtsiefer, eprint 0807.0991 (2008); J. Řeháček, B.-G. Englert, and D. Kaszlikowski Phys. Rev. A **70**, 052321 (2004).
 - [13] M. Riebe, K. Kim, P. Schindler, T. Monz, P. O. Schmidt, T. K. Körber, W. Hänsel, H. Häffner, C. F. Roos, and R. Blatt, Phys. Rev. Lett. **97**, 220407 (2006).
 - [14] D. Jurafsky, and J. H. Martin, *Speech and Language Processing*, Prentice Hall (2009).
 - [15] P. A. Estévez, C. J. Figueroa, and K. Saito, Proc. IEEE IJCNN **5**, 2724 (2005).
 - [16] This set of input states, taken from Ref. [13], is just one of the many possible choices one can use in quantum process tomography. It is important to understand that this set is by *no means* sanctioned to be the “standard” set of input states. Rather, these are four states of the six projectors of the standard six-outcome POM, but **any** four of the six states will serve the purpose equally well.